Reviewer's report

Title: Analysis and classification of oncology activities on the way to workflow based single source documentation in clinical information systems

Version: 3
Date: 28 June 2015
Reviewer: Brigitte Seroussi

Reviewer's report:

Cancer documentation is usually multiple, because it serves different objectives. On the one side, cancer patient data are needed to feed data bases such as registries for epidemiological analysis, and on the other side cancer patient data are needed to describe the patient case and apply evidence-based clinical practice guidelines to manage her/him according to best practices. The authors aim at building a solution that answer the two issues. The aim of the work is thus to develop a single-source documentation solution for the management of cancer patients, the feeding of cancer registries, and the collection of information for quality insurance. The article proposes a 14-step method to compile distributed cancer documentation about 13 different cancer entities. Based on the principle of a process analysis, refined and completed by interviews with experts, the method provides few grouped clinical pathways organized in 3 general categories of cancer management.

The article is well written but the English language should be improved. The sentences are very long which makes the article difficult to read. A lot of typos should be corrected.

In addition, if the article is structured according to the regular Background/Methods/Results/Discussion/Conclusion layout, the sections are not so clear. As far as I am concerned, if the Methods section describes the whole 14-step method, the Results section proposes a focus on the steps 12 and 13 which is also part of the Methods section. In addition, a similarity/conformity coefficient is introduced in the Results section and should clearly belong to the Methods section. Finally, the last paragraph of the Results section should be in the Discussion.

On the basis of the article, it is difficult to evaluate the scientific contribution of the work. The method described is not new: compilation of documentation, first draft of workflows and pathways, semi-structured interviews with experts, completion of original workflows and pathways, last round with the experts, this is common. Then, a generalization process has been handled on the workflows to generalize the common steps and build branching derivations to take into account particular cases, to finally get 3 general classes. This is also very common. New contributions as far as the methods are concerned have not been described. In addition, we don’t understand why the generalization stopped at the level of 3 classes. It would have been possible to generalize more and get only one class.
The criterion used to stop the generalization is not explained. It is particularly important because, intuitively, one should have generalized the first two classes (« Solid entities with surgical therapy » and « Solid entities with surgical therapy and additional therapeutic activities ») and get only 2 classes: either “with and without surgery”, or similarly “solid and non-solid entities”.

Another issue concerning the work presented is the fact that it is justified by the need to have a documentation to manage cancer patients that seems to be customized to the practices of the EUH. In addition, clinicians are said to deploy great efforts in the “jungle of treatment” to access the information about the “right actions at the right time”. This should be argued since in the case of cancer management, the pace of information update is so fast that clinicians rather look for state of the art practices in Pubmed and ASCO publications instead of relying on a built-in documentation. Finally, if it is a jungle when considering all the sub-specialties of oncology, it is not the case for cancer specialists that follow the publication of successful protocols or the discovery of new markers or exams in their specialty.

Major Compulsory Revisions

In a revision of the article, it would be important for the authors to describe more precisely their contribution. If their contribution concerns the methods, they should describe more precisely the methods and explain how previous methods were not sufficient to reach their goal. If their contribution concerns the results, they should evaluate the results and describe the impact of the solution developed on clinicians and/or patients and/or the hospital organization. The authors should more explain how their work is useful.

Minor Essential Revisions

- letters for affiliations are not in the right order of authors
- some results given in the abstract are not found in the document ("130 figures, 94 tables and 23 tumour classifications, 12 follow-up tables")
- in the Conclusions of the abstract, "of a comprehensive cancer" is not clear
- lot of typos should be corrected ("experts societies focused single cancer", "we discovered that is was", "transfer" is not appropriate (lg 110), "were common" instead of "are common" lg 113, "concept" lg 120 should be "structure", "minor adaptations" lg 212, "for the diagnostic and therapeutic workflows lg 213-214, "semantic gap und to achieve" lg 280, "for example of bioinformatics" lg 286, etc ....)
- authors should say if the experts consulted in the first and second rounds of interviews were the same
- the 13 cancer entities should be labelled the same lg 130-131 and lg 190-191
- the calculation of lg 200 is not clear: 13 entities are grouped into 11 cancer types (instead of 10) if colon+rectum = colon-rectum and plasmocytoma+acute myeloid leukemia=one disease, this makes 13-2 = 11
- the same applies when prostate, kidney and bladder are merged, and when
mamma and cervix are merged
- similarity/conformity coefficient should be normalised otherwise the absolute threshold at 6 does not work for long sequences
- "The above identified three classes should be treated separately when implementing clinical documentation modules because they are so distinct in their course and order of inpatient and outpatient episodes", Ig 242-243 should be reworded to show moderation (otherwise, this has to be proven)

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests.